Match Prediction using Random Forest Classification in Defense of the Ancients 2(DOTA 2)

Patlolla, Akhil Reddy
Computer Science Dept.
Northern Illinois University
apatlolla@niu.edu

Abstract – In the game of DOTA2, the players of one team coordinate among themselves to fight against other teams. The factors which affect the game are hero picking, lane selection and item build along with farming in the neutral camps. The major attributes of any hero are strength, agility and intelligence. Here, our main goal is to predict a probable winner based on the game metrics and subsequently build a prediction model.

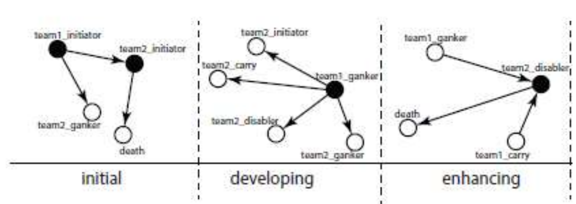
Keywords – DOTA2, MOBA, Multi-Player, Skill, Player Behavior Modeling, Random Forest, Skill, True skill, Area Under Curve.

I. Introduction

Multiple Online Battle Arena (MOBA) is a game-mode implementing real time strategy with two teams and five players each. A game-mode like this calls for immense coordination among the team members. An interesting part of the game is choosing a hero. There are around 113 heroes to choose from and five heroes per team need to be chosen. Hence, picking up the best team possible combination is highly crucial. Skills are shared or distributed among these heroes during the course of the game. An effective hero is picked among teammates who have parallel skill level. All these aspects involve heavy machine learning techniques.

II. Mechanics

In a MOBA game-mode, there are two types of combats: Farming and Ganking. Farming refers to the point of killing weaker units called Creeps. By killing these Creeps, players accumulate gold and experience. This enables the players buy items which help increase the attributes and special powers of their heroes as desired. Moving along the map and killing enemy is Ganking (gang-killing). Ganking is a better way than farming to gain experience and gold. This helps a team gain dominance in the game.



III. Hero Selection

In a team, there is a specific role assigned to each player.

Details & attributes of the roles :

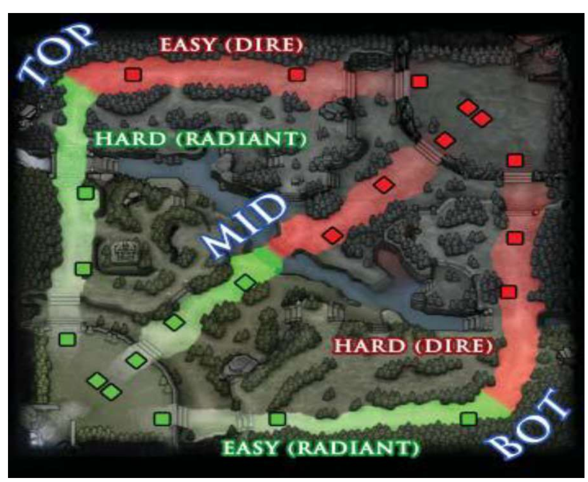
	Treant Protector	Phantom Lancer	Lina
Class	Strength	Agility	Intelligence
Strength	25	18	18
Agility	15	23	16
Intelligence	17	21	27
Game role	Durable Initiator Lane support Disabler	Carry Escape Pusher	Nuker Disabler Support

- **Carry** : A "Carry" game-role can do tremendous damage once they get a chance to develop. This means that they require protection early in the game but are responsible for most of the damage done eventually in the game.
- **Initiator :** Initiators are characters that usually begin large combats by doing large AOE (Area of Effect) or usually possess the abilities that can affect the positioning of the enemy team.
- **Disabler** : They have the ability to hinder the functionality of the enemy team by disabling or controlling them in some way.
- **Tank** : Tank has the ability to absorb copious amounts of damage. This enables the other members of the team to live longer in combat.
- **Ganker** : Ganker has the ability to deal with a large amount of damage quickly. Their main goal is to quickly kill enemies, so that combat ends as soon as possible.

There are various locations where players can gank depending on the location of the map and time.

IV. Winning strategies

There are locations where players play for farming and ganking. But, for the main game there are three main sectors in the map. The image below displays the details



Top, Mid, Bot are the zones in the map. The two sides of any zone are Dire and Radiant represented by red and green respectively. Mid has a balanced situation. All the squares in the map are towers. The left bottom corner and right top corner from the above picture are called ancient (a team's base / building) whole point of game which rotates around this to defend the ancient building.

V. Data Handling and Methodologies

The data was parsed from steam web API, which is a collection of game data. SORT operation is performed based on MMR (Match Making Rank) of the players and by selecting equal distribution of the players' data from the pool based on ranking. The selected data set comprises of 48 billion records. Task is to pick the features which affect the win prediction of the match. The game play involves metrics such as GPM (Gold Per Minute) i.e. average gold gained per every minute by each player, XPM (Experience Per Minute) average experience gained per every minute in the game, total gold earned in the game etc. Collected gold is spent towards in-game purchases such as items and buyback. Last hit is a metric which evaluates the player's skill to kill Creeps i.e. if a Creep is dead with the final hit of the player, then the respective hero will obtain bonus gold and experience. Also, last hit is recorded and higher the last hit, better chance of the hero to have better GPM, XPM.

Skills considered to build the Random Forest First Prediction Model:

'gold', 'gold_spent', 'gold_per_min', 'xp_per_min', 'kills', 'deaths','assists', 'denies', 'last_hits', 'stuns', 'hero_damage', 'hero_healing', 'tower_damage'

```
##
metrics.roc_auc_score(test_labels.values, test_probs[:,1])
0.50338291745162056

##
print(metrics.classification_report(test_labels.values, test_preds))
             precision    recall  f1-score   support

          0       0.48      0.46      0.47     48139
          1       0.52      0.55      0.53     51861

avg / total       0.50      0.50      0.50    100000
```

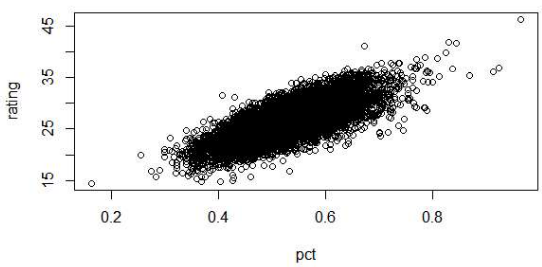
The AUC for this model is observed to be 0.50 which predicts that both the teams have equal probability to win. Hence, the model fails.

Considering the above model, correlation of all the features has been calculated with respect to Win Prediction.

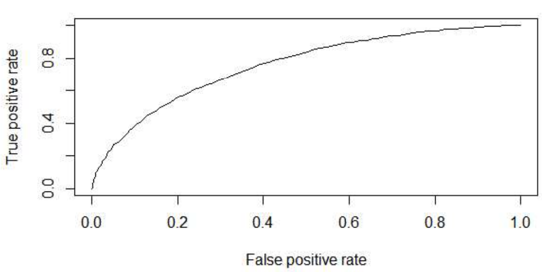
Feature	Correlation
gold	0.50
gold_spent	0.58
gold_per_min	**0.72**
xp_per_min	**0.69**
kills	0.61
deaths	0.28
assists	0.24
denies	0.12
last_hits	0.57

stuns	0.29
hero_damage	0.44
hero_healing	0.39
tower_damage	**0.71**

From the above table, gold_per_min, xp_per_min and tower_damage have higher correlation values. The average value of these three factors is known as 'True-Skill' and correlation of win prediction with this True-Skill is 0.75.



A prediction model is built considering True-Skill as feature over Random Forest Classifier and a test set of around 1 million records is supplied as test data for the model and the ROC curve for that prediction model is as follows:



The AUC for the above prediction is 0.75 which represents a better model than the prior one.

VI. Conclusion

The whole data set of matches is pulled form Valve's Steam web API which satisfies the levels of players of different level.

This prediction model can be used to read the game data, game statistics and predict the winning team.

References

[1] How Does He Saw Me? A Recommendation Engine for Picking Heroes in DOTA 2, Kevin Conley, Daniel Perry.

[2] The Well-Played MOBA: How DotA 2 and League of Legends use Dramatic Dynamics, Chris Winn.

[3] Modeling Intransitivity in Matchup and Comparison Data, Shuo Chen, Thorsten Joachims.

[4] Skill-Based Differences in Spatio-Temporal Team Behavior in Defense of The Ancients 2 (DOTA 2), Anders Drachen, Matthew Yancey, John Maguire, Derrek Chu, Iris Yuhui Wang, Tobias Mahlmann, Matthias Schubert, and Diego Klabajan.

[5] Identifying Patterns in Combat that are Predictive of Success in MOBA Games, David L. Roberts, Brent Harrison, Pu Yang.

[6] Calculating Optimal Jungling Routes in DOTA2 Using Neural Networks and Genetic Algorithms, Thomas E. Batsford.

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.